# EDGEWOOD CHEMICAL BIOLOGICAL CENTER

## U.S. ARMY RESEARCH, DEVELOPMENT AND ENGINEERING COMMAND
Aberdeen Proving Ground, MD  21010-5424

# A COMPARISON OF PREDICTIVE THERMO AND WATER SOLVATION PROPERTY PREDICTION TOOLS AND EXPERIMENTAL DATA FOR SELECTED TRADITIONAL CHEMICAL WARFARE AGENTS AND SIMULANTS II: COSMO-RS AND COSMOTHERM

Jerry B. Cabalo

RESEARCH AND TECHNOLOGY DIRECTORATE

April 2017

U.S. ARMY
RDECOM

Disclaimer

# REPORT DOCUMENTATION PAGE

| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From - To)* |
|---|---|---|
| XX-04-2017 | Final | Dec 2015 – Jun 2016 |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| A Comparison of Predictive Thermo and Water Solvation Property Prediction Tools and Experimental Data for Selected Traditional Chemical Warfare Agents and Simulants II: COSMO-RS and COSMOTherm | |
| | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER<br>CB10115 |

| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
|---|---|
| Cabalo, Jerry B. | |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |

| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) | 8. PERFORMING ORGANIZATION REPORT NUMBER |
|---|---|
| Director, ECBC, ATTN: RDCB-DRI-T, APG, MD 21010-5424 | ECBC-TR-1454 |

| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
|---|---|
| Defense Threat Reduction Agency, 8725 John J. Kingman Road, MSC 6201, Fort Belvoir, VA 22060-6201 | DTRA |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for public release: distribution unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT:**
The work described in this report is a follow-on to a previous study that compared the accuracy of a set of physico-chemical property prediction packages. In the previous study, we compared EPI Suite, ACD-Labs, and COSMO-RS software, as implemented in the ADF 2012 suite of programs for the physico-chemical properties of a set of traditional chemical warfare agents and selected simulants. To ensure comparison of the most up-to-date results, we repeated our COSMO-RS calculations with the implementation in the COSMOTherm suite of programs. The newer version utilizes the contribution from multiple conformers, a finer description of the molecular electronic structure, and a larger database of chemical properties to parameterize the program. We noted significant improvements in accuracy for all five physico-chemical properties. The resulting values were comparable in accuracy to those calculated using EPI Suite software.

**15. SUBJECT TERMS**

| | | | |
|---|---|---|---|
| Density functional theory | Electron density | Physical properties | Chemical properties |
| Theoretical prediction | Chemical agents | Simulants | |

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON<br>Renu B. Rastogi |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | UU | 32 | (410) 436-7545 |

Blank

# EXECUTIVE SUMMARY

The capability to accurately and reliably predict the physical and chemical properties of molecular compounds is highly desirable. In the case of industrial chemicals, the sheer number of possible compounds necessitates a predictive capability. Chemical and biological defense applications also benefit from a predictive capability, although there is a smaller number of potential chemical warfare agents. The need exists because only a few laboratories are capable of working with these compounds. Given this need, we performed a comparative study of the accuracy of a number of predictive software packages using a set of traditional chemical warfare agents, simulants, and a controlled substance. The results were published in a previous report (ECBC-TR-1259). However, in the previous study we utilized an older implementation of the COSMO-RS software (embedded in ADF 2012). To ensure the best comparison, we repeated our calculations using the newest approach as implemented in COSMOTherm.

There are two important differences between the newer implementation and the older version used in the previous report. First, a larger basis set was used to build up the description of the electronic structure of target molecules. Presumably, this leads to a more accurate description of the electronic structure of the molecules. Second, the newer implementation utilizes the contributions from multiple molecular conformations that are accessible at typical ambient temperatures. Because of these differences, we repeated the calculations using the newer implementation. We utilized the conformer generator (the COSMOconf) included with the COSMOTherm and then optimized the resulting structures using density functional theory at the BP86/TZVPD-Fine level of theory. The output of the calculations consisted of descriptions of the molecular surface charge for a set of conformations that we would expect to see at ambient temperature. We then directed these descriptions of each molecule to COSMOTherm to calculate boiling point, vapor pressure, water solubility, octanol/water partition coefficient ($pK_{ow}$), and the first hydrogen dissociation constant ($pK_a$). These results are compared to the results from ADF-COSMO-RS and EPI Suite software reported in the previous study. For all five physico-chemical properties, there were significant improvements in accuracy when the latest implementation of COSMO-RS (COSMOTherm) was used. The predicted values from COSMOTherm also proved comparable to those obtained from EPI Suite software.

Blank

# PREFACE

## Acknowledgments

Blank

# CONTENTS

# FIGURES

# TABLES

# A COMPARISON OF PREDICTIVE THERMO AND WATER SOLVATION PROPERTY PREDICTION TOOLS AND EXPERIMENTAL DATA FOR SELECTED TRADITIONAL CHEMICAL WARFARE AGENTS AND SIMULANTS II: COSMO-RS AND COSMOTHERM

## 1.        INTRODUCTION

### 1.1        Motivation

The purpose of this report is to present a comparison of experimental measurements to predictions from the most recent implementation of COSMOTherm (also known as COnductor-like Screening MOdel for Real Solvents [COSMO-RS]).[1-3]  The purpose of our original comparison was to provide government researchers with a basis to select predictive in-silico tools for physico-chemical properties of chemical warfare agents.  Given that we did not use the most recent version of COSMO-RS, it was necessary to repeat some of the calculations with updated results.

### 1.2        Background

The capability to predict the physical and chemical properties of chemical warfare agents is critical for the development of detection methods including forensics, countermeasures, and understanding the fate of these compounds in the environment.[4-7]  By their very nature, these chemical compounds are highly toxic and laboratories with the capability to work with these materials are limited.  As a result, it is not possible to characterize every compound of interest with respect to an unlimited number of properties.  The capability to reliably predict these properties for a wide range of compounds can also extend existing laboratory measurements.  A similar problem confronts regulatory agencies such as the Environmental Protection Agency (EPA), where thousands of different industrial chemical compounds are produced, yet there are insufficient resources to characterize potential toxicity of these compounds.  Fortunately, a number of different approaches to make reliable predictions are available.

Much progress in the prediction of physico-chemical properties has been made since it was first noted that the boiling point of hydrocarbons could be predicted by performing a regression on the number of carbon atoms in the molecule.[8]  In a previous report, Cabalo and Knox[9] used a number of available prediction software prediction packages on a selected set of chemical warfare agents and selected simulants, and compared the results.  A number of these software packages utilize group contribution methods, such as EPI Suite or ACD Labs.  For these methods, statistical regression of property values of a large training set of compounds is performed against a variety of chemical descriptors. Various chemical functional groups or unique groups of atoms serve as chemical descriptors.  As noted in the previous report, however, there are limitations to these approaches.  First, if a given element, functional group or molecular substructure in a target molecule is not well represented in the original training set, inaccurate predictions can result.  Second, it can be difficult for group contribution methods to account for

molecular symmetry. Third, the presence of structures that are best described by resonance structures can confuse the group contribution methods. Lastly, the presence of salt structures can also confuse group contribution methods. For this reason, inclusion of electronic structure methods such as density functional theory (DFT) have been considered for prediction of molecular properties.

Use of regression methods are still advantageous, yet many of the shortcomings of standard group contribution methods can be overcome using descriptors from electronic calculations rather than groupings within the chemical formula of a compound. In addition, the use of regression of empirical data in combination with electronic structure calculations has a number of advantages over calculations of properties totally from first principles. This is mainly due to the fact that for all but a few chemical compounds, a great deal of assumptions must be made to tractably calculate the electronic structure of many of the compounds of interest. To make predictions completely from first principles more difficult, different assumptions may be required for different classes of compounds. Thus, it is difficult to make reliable predictions based solely on first principles calculations. However, descriptors that go into a regression model can be taken from simple quantum mechanical calculations, e.g. dipole moment, or charge distribution.

There are a number of specific disadvantages that are overcome when utilizing descriptors calculated from first principles. First, relying on electronic structure calculations instead of group contributions with a multiplicity of atomic elements and chemical functional groups, greatly simplifies the relationship between the regression model and the descriptors. This simplification can greatly generalize the model training set so that compounds containing unusual elements such as arsenic, or unusual chemical linkages such as P-N-C in tabun, can be handled by the regression model. As a result, we expect the approach utilized in COnductor-like Screening MOdel-Real Solvents (COSMO-RS),[1; 2; 10; 11] to produce predictions that come the closest to the experimental values. Instead, in the previous study, the predicted values from the standard group contribution methods such as EPI Suite or ACD Labs more consistently approached the experimental values.

There are several reasons that can account for the larger root mean square error (RMSE) values obtained while using COSMO-RS. In our original report we hypothesized that the size of the training set (the COSMO-RS version we utilized was had a parameterization with 642 compounds) affected the accuracy. In response to the previous report we published, the authors of COSMO-RS informed us that our study utilized a much older version, and that a number of improvements had been made that could significantly improve the results from COSMO-RS. These improvements addressed a number of other issues in addition to parameterization that could affect the accuracy of the COSMO-RS predictions. First, parameterization has been done with an improved basis set, going from the triple zeta TZP basis set[12] to the def2-TZVPD basis set[13] that includes more polarization functions. Additional polarization functions permit more accuracy with respect to heavier elements. Also, the grid used to calculate the COSMO screening charge is finer compared to previous implementations. Lastly, and possibly most importantly, the newer implementation takes into account thermally accessible molecular conformations. The charge distribution on the COSMO cavity surface can

change significantly with conformation.  Therefore, we expect improvements in accuracy if the calculation better accounts for the physics in the real system.

## 2.        EXPERIMENTAL PROCEDURE

We repeated the calculations for the set of traditional chemical warfare agents and simulants that had been done with the implementation of COSMO-RS in the Amsterdam Density Functional Code (ADF 2012)[14] for boiling point, vapor pressure, water solubility, octanol/water partition coefficient, and the dissociation constant for the first proton $pK_a$. However, we utilized the more up to date procedure.  First, a set of conformer structures were determined using the COSMOconf program.[15]  This routine automatically generates a set of conformer structures including the charge distribution on the COSMO cavity surface.  Then, using the Turbomole[16] suite of programs, the resulting structures were geometry optimized using the BP86 density functional with the triple zeta def2-TZVPD basis set.  Geometry optimized conformers for the gas phase were also generated but without the COSMO polarizable continuum model.  For $pK_a$ calculations, it was necessary to also calculate the optimized conformer structures for the ionic forms for VG, VX, Red9, GA (tabun), glycerol, metamidophos, and malathion.  For calculations involving one (water solubility) or two solvents (octanol/water partition), the conformer structures and charge distribution data of water and octanol were obtained from precalculated files.  For each compound, a set of conformers resulted.  For each conformer, there was a *.cosmo and *.energy file.  These files were in turn used in the COSMOTherm calculation.  Lastly, the BP_TZVPD_FINE_C30_1501.ctd database was used for all predictions.

## 3.        RESULTS AND DISCUSSION

### 3.1        Boiling Points

Table 1 compares the predicted values obtained from the implementation of COSMO-RS that is embedded with the Amsterdam Density Functional (ADF) code.  The results of EPI Suite calculations are also included for reference.  Generally, the newer implementation of COSMO-RS resulted in significant reduction of error compared to the experimental values. For the organophosphate compounds and mustard agents, e.g. DMMP or VX, the COSMO-RS consistently overestimates the boiling point by an average of ~$100^o$ C.  The COSMOTherm results reduce that error on average to ~$32^o$ C.  Some of the more drastic corrections occur for VX or DMMP with values of over $200^o$ C. However, for GA, GB, GD, and GF, the correction factor from the newest implementation of COSMO-RS (COSMOTherm) is quite moderate.  The COSMOTherm calculations for DMMP, GA, GB, GD, GF, and HD utilized 3, 2, 3, 2, 7, and 4 conformer structures, respectively.  DMMP and HD experienced significant improvement with COSMOTherm relative to the implementation in ADF.  However, GA, GB, and GF experience little change in value when going to the newer implementation.  The chief differences between the G agents and other organophosphate compounds is the presence of unusual linkages to the phosphorus atom.  In GA, there is a cyanide functional group and a tertiary amine attached to the phosphorus.  In GB, GD, and GF, there is a fluorine atom attached to the phosphorus atom.  Yet, in contrast to GB, GD and GF have a more significant organic component.  It would appear that

as more conventional portions of the molecule contribute more to a given property, these can balance out the "unusual" portion. For the sulfur and nitrogen mustard agents, the COSMO-therm results are very close to the experimental value.

For all cases, except the Red 9 dye molecule, the COSMOTherm predicted lower boiling points than that obtained with ADF COSMO-RS. In most cases, the reduction in predicted boiling point approached the experimental value. For Lewisite L1, we actually see COSMOTherm underpredict the boiling point. We do not expect differences between the results from COSMOTherm and the older version of COSMO-RS to arise from the set of compounds used for parameterization. We do not expect the increased difference between the experimental boiling point value and the predicted value to arise from the basis set used to describe the electronic structure. For the first set of calculations we used the TZP basis set. For the newer calculations we used the larger TZVPD split valence basis set with polarization functions. The number of compounds in the database used for parameterization has since been added to, and we would expect the accuracy to increase. However, we do note that COSMOTherm utilizes a number of thermally accessible conformations to determine the boiling point. The implementation of COSMO-RS within ADF in our first set of calculations utilized only one conformation. We conjecture that the predicted boiling point value from the ADF-COSMO-RS was close to the experimental value by coincidence. A similar situation arises for glycerol. The value reported in the table corresponds to the result using the conformational structures we calculated with COSMOconf. Using the precalculated COSMO files for the conformations of glycerol, we obtained a much closer value to experiment. We also expect that the conformation generator cannot sample all of the potential energy surface, and that the structures/COSMO files we generated are not necessarily optimal. However, because predicted values from COSMOTherm typically depend on a weighted average of multiple conformers, we expect COSMOTherm to be more consistent and less dependent on the starting geometry used in the calculation. With respect to comparison to the accuracy EPI Suite results, the COSMOTherm results give roughly similar results for the set of compounds investigated.



DIMP          DMMP          GA
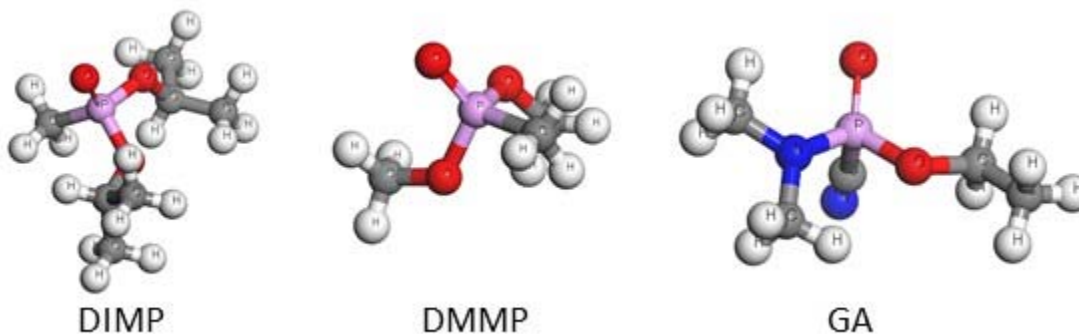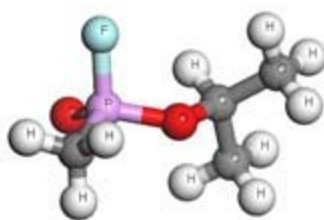
GB
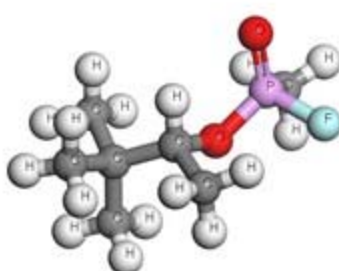
GD

GF
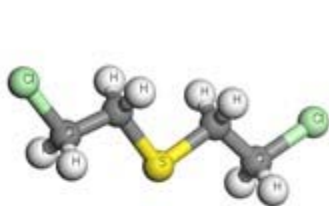
HD

HN1
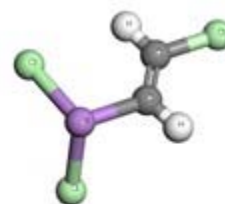
L1
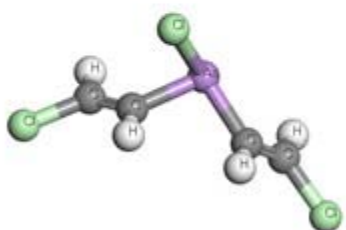
L2
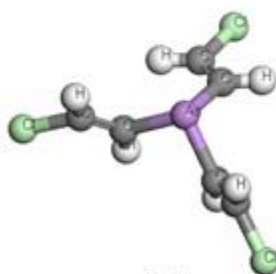
L3
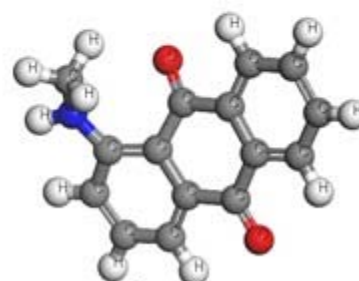
Red 9
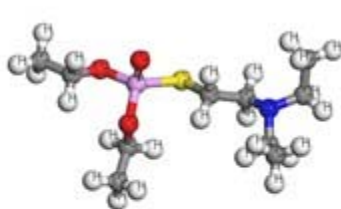
VG
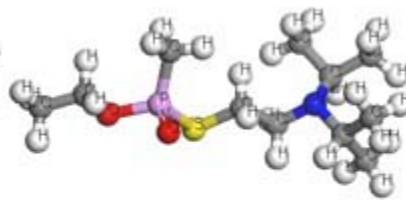
VX
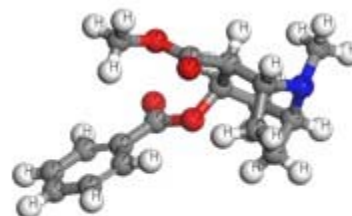
Cocaine
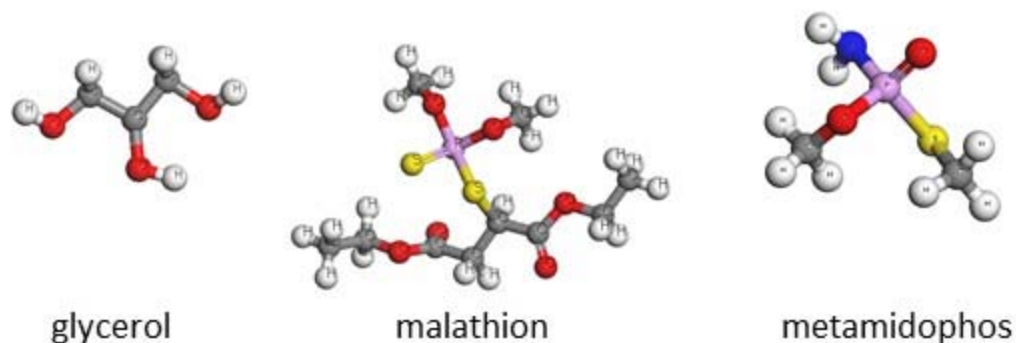
**Figure 1.** Molecular structures of the compounds utilized in this study.

**Table 1.** Comparison of Predicted Boiling Points from COSMOtherm and COSMO-RS as Implemented in ADF 2012[17-24]

| Compound | Boiling Point ($^{o}$C) | | | |
| | EPI Suite | COSMOtherm (TZVPD-Fine) | ADF COSMO-RS | Exp |
|---|---|---|---|---|
| DIMP | 210 | 238 | 283 | *high |
| DMMP | 152 | 207 | 394 | 181 |
| GA | 267 | 307 | 324 | 240 |
| GB | 140 | 195 | 218 | 147 |
| GD | 183 | 230 | 306 | 198 |
| GF | 223 | 265 | 278 | 239 |
| HD | 210 | 219 | 265 | 216 |
| HN1 | 212 | 207 | 301 | 194 |
| L1 | 156 | 164 | 215 | 196.6 |
| L2 | 204 | 244 | 290 | n/a |
| L3 | 247 | 244 | 318 | n/a |
| Red9 | 397 | 427 | 382 | n/a |
| VG | 337 | 345 | 613 | n/a |
| VX | 321 | 334 | 550 | 298 |
| cocaine | 363 | 452 | 505 | n/a |
| glycerol | 231 | 222 | 325 | 290 |
| malathion | 351 | 434 | 701 | *high |
| metamidophos | 223 | 322 | 324 | *high |

*high: the literature reports no value but the term "high".

## 3.2      Vapor Pressures

To generalize the results in Table 2 for the comparison of COSMOTherm, and ADF-COSMO-RS, we see a contribution to the accuracy from the inclusion of multiple molecular conformations in the calculation of vapor pressure.  For rigid molecules with limited degrees of freedom, we see good agreement between the ADF COSMO-RS and COSMOTherm results.  GA, GB, HD, HN1, L1, and Red 9, have rigid structures, or equivalent conformations. As a result, predicted values from COSMOTherm or ADF-COSMO-RS are either similar to each other, or have similar differences with experiment.  For GA, GB, HN1, and L1, both ADF-COSMO-RS and COSMOTherm are within an order of magnitude to the experimental value. However, for Red 9, both COSMOTherm and ADF-COSMO-RS are nearly an order of magnitude off with COSMOTherm an order of magnitude too small, and ADF-COSMO-RS  an order of magnitude too high.  We expect that because Red 9 is a crystalline solid, the primary cause for the discrepancy between experiment and theory is the determination of the heat of fusion.

**Table 2.** Comparison of Vapor Pressure Values from COSMOTherm, ADF-COSMO-RS, and Experiment (see references for Table 1)

|  | Vapor Pressure (Pa) | | | |
|---|---|---|---|---|
| Compound | EPI Suite | COSMOtherm (TZVPD-Fine) | ADF COSMO-RS | Exp |
| DIMP | 3.00E+01 | 1.72E+01 | 4.60E-01 | 3.70E+00 |
| DMMP | 1.20E+02 | 1.03E+02 | 1.30E+01 | 1.30E+02 |
| GA | 6.20E+00 | 2.25E+00 | 3.30E+00 | 9.30E+00 |
| GB | 6.10E+02 | 1.64E+02 | 1.40E+02 | 3.80E+02 |
| GD | 5.30E+01 | 2.70E+01 | 9.20E+00 | 5.30E+01 |
| GF | 6.50E+01 | 6.59E+00 | 2.30E+01 | 5.90E+00 |
| HD | 2.10E+01 | 2.68E+01 | 3.00E+01 | 1.50E+01 |
| HN1 | 2.60E+01 | 5.06E+01 | 1.30E+01 | 3.30E+01 |
| L1 | 7.90E+01 | 2.27E+02 | 2.00E+02 | 7.70E+01 |
| L2 | 3.90E+01 | 3.77E+00 | 1.60E+01 |  |
| L3 | 3.60E+01 | 3.69E+00 | 8.80E+00 |  |
| Red9 | 4.10E-05 | 1.49E-03 | 7.10E-01 | 9.30E-02 |
| VG | 3.60E-02 | 5.37E-02 | 5.00E+03 | 3.50E-02 |
| VX | 2.90E-01 | 1.14E-01 | 4.00E-03 | 9.30E-02 |
| cocaine | 1.70E-03 | 1.06E-03 | 1.20E-02 | 3.90E-05 |
| glycerol | 1.10E-02 | 2.10E+00 | 1.00E-03 | 2.20E-02 |
| malathion | 1.70E-02 | 7.19E-04 | 0.00E+00 | 4.50E-04 |
| metamidophos | 9.10E+00 | 1.73E-01 | 2.60E-02 | 4.70E-03 |

For larger molecules with significant hydrocarbon components with many degrees of freedom and several conformations, such as DIMP, GD, GF, VG, VX, cocaine, and malathion, we see orders of magnitude improvement in the agreement between prediction and experimental measurement. The most dramatic improvement in accuracy is for VG, where the vapor pressure from ADF-COSMO-RS is much higher (5000 Pa) than both the COSMOTherm predicted value ($5.37 \times 10^{-2}$ Pa) and the experimental value ($3.50 \times 10^{-2}$ Pa). For the set of compounds considered, especially with a significant hydrocarbon component with a selection of available conformations, COSMOTherm makes a significant improvement. In the case of metamidophos, although both the COSMOTherm and ADF-COSMO-RS predicted values are both small, less than 1 Pa, we note that the ADF-COSMO-RS value is an order of magnitude closer in value to experiment than the COSMOTherm value. We conjecture that the value predicted from ADF-COSMO-RS is accidentally close to the experimental value. Given that only one conformation was used to determine the prediction from ADF-COSMO-RS, it is possible that that single conformation gave rise to a lower vapor pressure value. A systematic study comparing the results for vapor pressure from different individual conformations as well as various combinations would definitively determine if that result is accidental, but is not done here due to constraints in resources.

When we consider the EPI Suite results, we see very similar results with COSMOTherm. Some notable exceptions are GF, glycerol, malathion, and metamidophos. For the organophosphate compounds, the COSMOTherm results are closer to the experimental value. For glycerol, the prediction from EPI Suite is closer to the true value. Based on this set of results, COSMOTherm has similar accuracy to EPI Suite.

## 3.3 Water Solubility

We generally see an improvement in accuracy when comparing water solubility predictions from COSMOTherm to COSMO-RS. Table 3 shows the numerical results of the study. For low solubility to moderate solubility compounds we see the greatest improvement in agreement between predicted water solubility and the experimental measurements. ADF-COSMO-RS underpredicts the solubility by two orders of magnitude for VG and VX in comparison to COSMOTherm. For Red 9, COSMOTherm reduces the predicted vapor pressure almost two orders of magnitude, although it is still almost two orders of magnitude greater than the experimental value. We also see moderate, within an order of magnitude, improvements in accuracy for GA, GD, GF, and L1. For highly soluble compounds with a high affinity for water, we do not see much difference between the values from ADF-COSMO-RS, COSMOTherm, and experimental values.

**Table 3.** Comparison of Predicted Water Solubilities (mg/L) from COSMOTherm and ADF COSMO-RS 2012[25-29]

| Compound | Water Solubility (mg/L) | | | |
|---|---|---|---|---|
| | EPI Suite | COSMOtherm (TZVPD-Fine) | ADF COSMO-RS | Exp |
| DIMP | 7.30E+03 | 1.35E+05 | 1.00E+06 | 1.50E+03 |
| DMMP | 3.20E+05 | 1.00E+06 | 1.00E+06 | 1.00E+06 |
| GA | 3.20E+04 | 3.27E+04 | 1.00E+04 | 9.80E+04 |
| GB | 4.60E+04 | 8.27E+05 | 1.00E+06 | 1.00E+06 |
| GD | 1.60E+03 | 1.56E+04 | 1.00E+04 | 2.00E+04 |
| GF | 2.10E+03 | 1.47E+04 | 1.30E+04 | 3.70E+03 |
| HD | 6.10E+02 | 1.35E+03 | 3.70E+02 | 6.80E+02 |
| HN1 | 4.00E+04 | 2.58E+04 | 1.30E+02 | 1.60E+02 |
| L1 | 2.60E+02 | 3.23E+02 | 1.80E+03 | 5.00E+02 |
| L2 | 2.90E+01 | 1.88E+01 | 2.00E+02 | |
| L3 | 3.30E+00 | 1.86E+01 | 1.90E+01 | |
| Red9 | 6.80E-01 | 6.38E+00 | 1.70E+02 | 1.20E-01 |
| VG | 6.60E+03 | 3.86E+04 | 1.00E+02 | 3.00E+04 |
| VX | 3.20E+03 | 5.74E+04 | 5.00E+02 | 3.00E+04 |
| cocaine | 1.30E+03 | 9.64E+03 | 3.50E+02 | 1.80E+03 |
| glycerol | 1.00E+06 | 1.00E+06 | 1.00E+06 | 1.00E+06 |
| malathion | 7.80E+01 | 2.17E+01 | 6.60E+01 | 1.40E+02 |
| metamidophos | 4.00E+05 | 1.00E+02 | 1.00E+06 | 1.00E+06 |

We attribute the improvement in accuracy of COSMOTherm to the parameterization of the larger basis set TZVPD, where we expect a more accurate and finer grained representation of the charge distribution on the surface of the molecule. If the contribution from conformers were more important, then we would expect the predicted water solubility values from ADF-COSMO-RS and COSMOTherm for Red 9 to not have much difference since Red 9 is a rigid molecule with limited degrees of freedom. On the other hand, for molecules that have significant hydrocarbon side chains with many internal degrees of freedom of motion, we would expect significant differences between the predictions of ADF-COSMO-RS and COSMOTherm. However, the values predicted for DIMP, GA, GD, GF, HD, LN1, are all quite close, even though DIMP, GA, GD, and GF have a number of conformers that contribute to the predicted solubility value. For compounds similar to the set of compounds shown in Table 3, we conclude accurate representation of surface charge on the molecule is more critical than the number of conformers. It is likely that the affinity for water does not change much with conformation.

We do acknowledge that for three compounds, HN1, malathion, and metamidophos, the result from COSMOTherm is not as close to the experimental data as that for ADF-COSMO-RS. For malathion, the discrepancy is negligible, given that for both COSMOTherm and ADF-COSMO-RS the predicted solubility values are within an order of

magnitude of the experimental value. HN1 does not possess strange molecular linkages, yet the result we obtained with COSMOTherm overpredicts the water solubility by two orders of magnitude. The discrepancy is even worse for metamidophos, where the water solubility is underestimated by four orders of magnitude. It can be seen in Figure 1 that these two compounds have amine groups in common. Yet, other compounds such as GA, Red 9, VG, VX, and cocaine possess amine groups and produce very good predictions for solubility. We can only conjecture that the error for metamidophos arises from the unusual phosphorus-nitrogen linkages.

When comparing the accuracy of the COSMOTherm predictions to that of EPI Suite, we see roughly equivalent performance. Of the 17 compounds compared, COSMOTherm had 10 compounds that had smaller errors than the EPI Suite predictions, although for the most part the differences between the two predictions were similar. Both EPI Suite and COSMOTherm appear to have a similar number of compounds (although not the exact same ones) that have significant differences with the experimental measurements.

## 3.4       Octanol/Water Partition Coefficients ($pK_{ow}$)

The predicted values for the octanol/water partition coefficient ($pK_{ow}$) from ADF-COSMO-RS and COSMOTherm are compared to the experimental values in Table 4. In most cases, the value predicted by COSMOTherm had a smaller $pK_{ow}$ value compared to ADF-COSMO-RS, except for the Lewisites (L1, L2, and L3) and malathion. In terms of closeness to the experimental values, we see consistent improvement for COSMOTherm. The ADF-COSMO-RS predictions of $pK_{ow}$ values are more than a whole pK unit away from experiment, such as for GA, HD, HN1, Red 9, VG, VX, cocaine, and malathion. In contrast, the differences between prediction and measurement are reduced to less than 1 pK unit with COSMOTherm, except for VG and malathion. We saw from section 3.3, that for water solubility alone, for some compounds, we obtained better results with ADF-COSMO-RS. However, with COSMOTherm, the improvement is consistent. Because the $pK_{ow}$ depends on the ratio of octanol to water solubilities, we expect any systematic errors present in the COSMOTherm calculation to cancel.

We do not expect that the inclusion of contributions from multiple conformers greatly contributed to the accuracy of the COSMOTherm $pK_{ow}$ values. We see consistent improvement in accuracy whether the target molecule has limited degrees of freedom, or if the molecule has many degrees of freedom. We see improvement in the COSMOTherm predictions for Red 9 as well as for VG and VX. We attribute the improvement to the improved depiction of the electronic structure of the target.

**Table 4.** Comparison of Octanol/Water Partition Coefficients from COSMOTherm, ADF COSMO-RS, and Experiment

| | $pK_{ow}$ (Octanol/water partition) | | | |
|---|---|---|---|---|
| Compound | EPI Suite | COSMOtherm (TZVPD-Fine) | ADF COSMO-RS | Exp |
| DIMP | 1.2 | 1.30 | 1.40 | 1.03 |
| DMMP | -0.6 | -0.55 | -0.50 | -0.61 |
| GA | 0.3 | 0.72 | 1.60 | 0.38 |
| GB | 0.3 | 0.59 | 1.10 | 0.30 |
| GD | 1.7 | 1.78 | 2.30 | 1.78 |
| GF | 1.6 | 1.59 | 2.20 | n/a |
| HD | 2.4 | 2.29 | 2.80 | 1.37 |
| HN1 | 1.4 | 2.91 | 3.60 | 2.02 |
| L1 | 2.6 | 2.86 | 2.70 | 2.56 |
| L2 | 3.5 | 4.42 | 3.50 | |
| L3 | 4.5 | 4.43 | 4.50 | |
| Red9 | 4.1 | 3.09 | 3.60 | 4.10 |
| VG | 1.7 | 3.50 | 4.80 | 1.70 |
| VX | 2.1 | 3.08 | 4.20 | 2.09 |
| cocaine | 2.2 | 3.00 | 3.90 | 2.30 |
| glycerol | -1.7 | -1.66 | -1.70 | -1.76 |
| malathion | 2.3 | 3.99 | 3.60 | 2.36 |
| metamidophos | -0.9 | -0.39 | 0.10 | -0.80 |

When considering the partition coefficient predictions from EPI Suite, we see for the most part excellent agreement between the EPI Suite and COSMOTherm calculations. Some exceptions are VG, VX, and malathion. These compounds differ from the experimental values by a pK unit or more. This is surprising since both EPI Suite and COSMOTherm have good agreement with respect to water solubility. To determine the reason for this discrepancy, we recommend future work to investigate how individual contributions from different conformations of octanol affect the computed value of $pK_{ow}$.

**3.5 Comparison of Dissociation Constants ($pK_a$) Values from COSMOTherm and ADF COSMO-RS**

Table 5 shows the results from calculations of $pK_a$ using COSMOTherm and the implementation of COSMO-RS in ADF. For the set of compounds chosen in this study, there is unfortunately limited data in the literature (this is not surprising given the controlled nature and toxicity of these compounds). We were able to locate experimental values of the first $pK_a$ for HN1, cocaine, VX, and glycerol. Both ADF-COSMO-RS and COSMOTherm performed quite well for these measurement, with difference from experimental values typically less than 1 pK unit. We see improvement in accuracy with COSMOTherm for HN1 and VX, but not for

glycerol.  Although further study is necessary to determine the reason for this, we observe that the dissociation constant for HN1 and VX are protonated amines rather than neutral alcohols.  Perhaps accuracy is better for some classes of compounds than others.  Based on this data set we expect gains in accuracy for $pK_a$ predictions when using the newest implementation of the method to be found in COSMOTherm.

**Table 5.** Comparison of pKa Values from Experiment, and Predictions from COSMOTherm and COSMO-RS

| Compound | pK$_a$ (1st proton) COSMOtherm (TZVPD-Fine) | ADF COSMO-RS | Exp |
|---|---|---|---|
| DIMP | | | |
| DMMP | | | |
| GA | -3.77 | -2.4 | |
| GB | | | |
| GD | | | |
| GF | | | |
| HD | | | |
| HN1 | 6.81 | 7.40 | 6.57 |
| L1 | | | |
| L2 | | | |
| L3 | | | |
| Red9 | -0.16 | 2.20 | |
| VG | 8.35 | 6.20 | |
| VX | 8.87 | 7.90 | 8.60 |
| cocaine | | 9.90 | 8.90 |
| glycerol | 11.38 | 12.30 | 14.15 |
| malathion | | | |
| metamidophos | | 3.20 | n/a |

## 4.        CONCLUSIONS

We repeated calculations on the set of compounds we reported previously using the newest implementation of COSMO-RS, as found in the COSMOTherm software.  These were done for boiling point, vapor pressure, water solubility, water/octanol partitioning coefficient ($pK_{ow}$), and the first dissociation constant in water ($pK_a$). Significant improvements in accuracy relative to the older implementation from the previous study have been made for most predicted property values, although there are a number of exceptions.  We attribute the improvement to both the improved representation of the molecular electronic structure with a larger basis set, inclusion of contributions from multiple conformations, as well as the use of a larger set of compounds used in the parameterization. We noted that there was across the board improvement, most notably for boiling point and vapor pressure predictions.  This proved true whether the molecule of interest had many internal degrees of freedom with many possible

conformers, or if the molecule was rigid with limited conformational possibilities. However, improvement was most dramatic for molecules possessing a significant hydrocarbon component with many degrees of freedom. For such molecules, we expect the improvement in accuracy when using the newer method implementation in COSMOTherm to arise from the contribution of conformers. However, we also observed improvement in accuracy for molecules with limited conformation possibilities. Although for some compounds and properties there was an increase in the error when using the newer COSMOTherm method, we attribute that to coincidence. Because the older method employed in ADF-COSMO-RS utilizes single conformer structures, it is possible that a less stable structure could produce a big effect on a predicted value, and that by coincidence, it could be closer to the experimental value.

We also compared the COSMOTherm predicted property values to those calculated with EPI Suite that appeared in our earlier report. Where previously the discrepancy between the ADF-COSMO-RS and EPI Suite predicted values were significant, we now see similar agreement between the methods. Given that the two methods utilize different approaches to make predictions, it may be possible to use both methods in tandem for greater reliability.

Blank

# LITERATURE CITED

[1]     Klamt, A., Jonas, V., Buerger, T., Lohrenz, J. C. Refinement and Parameterization of COSMO-RS. *Journal of Physical Chemistry A,* **1998***,* 102, 5074-5085, doi:10.1021/jp980017s.

[2]     Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.,* **2011***,* 1, 699-709, doi:10.1002/wcms.56.

[3]     Klamt, A. COSMO-RS for aqueous solvation and interfaces. *Fluid Phase Equilib.,* **2016***,* 407, 152-158, doi:10.1016/j.fluid.2015.05.027.

[4]     Carlsen, L. A QSAR approach to physico-chemical data for organophosphates with special focus on known and potential nerve agents. *Internet Electron. J. Mol. Des.,* **2005***,* 4, 355-366.

[5]     Lill, M. A., Dobler, M., Vedani, A. In silico prediction of receptor-mediated environmental toxic phenomena - Application to endocrine disruption. *SAR and QSAR in Environmental Research,* **2005***,* 16, 149-169, doi:10.1080/10629360412331319826.

[6]     Cronin, M. T. D. The Role of Hydrophobicity in Toxicity Prediction. *Current Computer-Aided Drug Design,* **2006***,* 2, 405-413.

[7]     Cui, X., Wang, Z., Zhang, Z., Yuan, X., Harrington, P. d. B. QSAR Study on the Toxicity of Phenols for Fathead Minnows by Using Support Vector Machine and Neural Networks, 2008.

[8]     Vernon, H. M. Molecular weights of liquids as evinced by their boiling points. *Chemical News and Journal of Industrial Science,* **1892***,* 64, 54-58.

[9]     Cabalo, J., Knox, C. K., 2014. A Comparison of QSAR Based Thermo and Water Solvation Property Prediction Tools and Experimental Data for Selected Traditional Chemical Warfare Agents and Simulants, ECBC-TR-1259, pp. 36, 2014.

[10]    Freire, M. G., Santos, L. M. N. B. F., Marrucho, I. M., Coutinho, J. A. P., 2010. Predicting the thermodynamic behaviour of water + ionic liquids systems using COSMO-RS. John Wiley & Sons, Inc., pp. 101-121, 2010.

[11]    Klamt, A. COSMO-RSC: Second-Order Quasi-Chemical Theory Recovering Local Surface Correlation Effects. *J Phys Chem A,* **2016***,* 120, 2049-56.

[12]    Barbieri, P. L., Fantin, P. A., Jorge, F. E. Gaussian Basis Sets of Triple and Quadruple Zeta Valence Quality for Correlated Wave Functions. *Molecular Physics,* **2006***,* 104, 2945.

[13]    Rappoport, D., Furche, F. Property-optimized Gaussian Basis Sets for Molecular Response Calculations. *Journal of Chemical PHysics,* **2010***,* 133, 134105.

[14]     van Gisbergen, S. J. A., Snijders, J. G., Baerends, E. J. Implementation of time-dependent density funcitonal response. *Computer Physics Communications,* **1999***,* 118, 119-138, doi:10.1016/S0010-4655(99)00187-3.

[15]     COSMOconf 4.0, 2015, available from http://www.cosmologic.de.

[16]     Turbomole 7.02 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, GmbH, T.; available from http://www.turbomole.com.

[17]     Rosenblatt, D. H., Miller, T. A., Dacre, J. C., Muul, I., Cogley, D. R., 1975. Problem Definition Studies of Potential Environmental Pollutants.  II. Physical, Chemical, and Toxicological and Biological Properties of 16 Substances. U. S. Army Medical Bioengineering Research and Development Laboratory, Fort Detrick, MD, 1975.

[18]     Boublik, T., Fried, V., Hala, E. Physical Sciences Data, Vol. 17: The Vapor Pressures of Pure Substances. 2nd Ed. Elsevier, 1984.

[19]     Munro, N. B., Talmage, S. S., Griffin, G. D., Waters, L. C., Watson, A. P., King, J. F., Hauschild, V. The sources, fate, and toxicity of chemical warfare agent degradation products. *Environmental Health Perspectives,* **1999***,* 107, 933-974.

[20]     Baughman, G. L., Weber, E. J. Estimation of water solubility and octanol/water partition coefficient of hydrophobic dyes. Part I. Relationship between solubility and partition coefficient. *Dyes Pigm.,* **1991***,* 16, 261-71, doi:10.1016/0143-7208(91)85016-2.

[21]     Lawrence, A. H., Elias, L., Authier-Martin, M. Determination of amphetamine, cocaine, and heroin vapor pressures using a dynamic gas blending system and gas chromatographic analysis. *Can. J. Chem.,* **1984***,* 62, 1886-8, doi:10.1139/v84-323.

[22]     Daubert, T. E., Bartakovits, R. Prediction of critical temperature and pressure of organic compounds by group contribution. *Ind. Eng. Chem. Res.,* **1989***,* 28, 638-41, doi:10.1021/ie00089a023.

[23]     USDA, USDA Pesticide Properties Database.

[24]     Tomlin, C. D. S., Editor. The Pesticide Manual, Eleventh Edition. Br Crop Prot Counc, 1997.

[25]     Krikorian, S. E., Chorn, T. A., King, J. W., Ellzy, M. W., 1987. Determination of the partition coefficients of organophosphorus compounds using high-performance liquid chromatography. Battelle Columbus Lab., pp. 21 pp., 1987.

[26]     Bennett, S. R., Bane, J. M., Benford, P. J., Pyatt, R. L., 1984. Environmental Hazrds of Chemical Agent Simulants. 1984.

[27]     Opresko, D. M., Young, R. A., Faust, R. A., Talmage, S. S., Watson, A. P., Ross, R. H., Davidson, K. A., King, J. Chemical warfare agents: estimating oral reference doses. *Reviews of environmental contamination and toxicology,* **1998***,* 156, 1-183.

[28]     Seidell, A. Solubilities of Organic Compounds. Van Nostrand, New York, NY, 1941.

[29]     Gupta, R. C. Handbook of Toxicology of Chemical Warfare Agents. Elsevier, New York, NY, 2009.

Blank

# ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| ADF | Amsterdam Density Functional Code |
| COSMO | COnductor-like Screening MOdel |
| COSMO-RS | COSMO for Real Solvents |
| DFT | Density Functional Theory |
| DTRA | Defense Threat Reduction Agency |
| DIMP | di-isopropyl methyl phosphonate |
| DMMP | di-methyl methyl phosphonate |
| ECBC | Edgewood Chemical Biological Center |
| GA | Tabun |
| GB | Sarin |
| GD | Soman |
| GF | Cyclosarin |
| HD | Distilled sulfur mustard |
| HN1 | Nitrogen mustard 1 |
| L1 | Lewisite 1 |
| L2 | Lewisite 2 |
| L3 | Lewisite 3 |
| VG | (see Figure 1) |
| VX | (see Figure 1) |

# DISTRIBUTION LIST

The following individuals and organizations were provided with one Adobe portable document format (pdf) electronic version of this report:

U.S. Army Edgewood Chemical
Biological Center (ECBC)
RDCB-DRI-T
ATTN: Cabalo, J.
      Goad, A.


Defense Threat Reduction Agency
J9-CBS
ATTN: Graziano, A.


Department of Homeland Security
RDCB-PI-CSAC
ATTN: Negron, A.
DHS-S&T-RDP-CSAC
ATTN: Strang, P.

G-3 History Office
U.S. Army RDECOM
ATTN: Smart, J.


ECBC Technical Library
RDCB-DRB-BL
ATTN: Foppiano, S.
      Stein, J.


Office of the Chief Counsel
AMSRD-CC
ATTN: Upchurch, V.


Defense Technical Information Center
ATTN: DTIC OA


ECBC Rock Island
RDCB-DES
ATTN: Lee, K.
RDCB-DEM
ATTN: Grodecki, J.

EDGEWOOD CHEMICAL BIOLOGICAL CENTER

A U.S. Army RDECOM Laboratory